Volume 13 • Number 8 • 2010
VALUE IN HEALTH

The Role of Observational Investigations in Comparative Effectiveness Research

Nicholas F. Marko, MD,[1] Robert J. Weil, MD[1,2]

[1]Department of Neurosurgery, The Neurological Institute, Cleveland Clinic, Cleveland, OH, USA; [2]Brain Tumor and Neuro-Oncology Center, Cleveland Clinic, Cleveland, OH, USA

ABSTRACT

Introduction: Comparative effectiveness research (CER) seeks to inform clinical decisions between alternate treatment strategies using data that reflects real patient populations and real-world clinical scenarios for the purpose of improving patient outcomes. There are multiple clinical situations where the unique characteristics of observational investigations can inform medical decision-making within the CER paradigm. Accordingly, it is critical for clinicians to appreciate the strengths and limitations of observational research, particularly as they apply to CER.

Methods: This review focuses on the role of observational research in CER. We discuss the concept of evidence hierarchies as they relate to observational research and CER, review the scope and nature of observational research, present the rationale for its inclusion in CER investigations, discuss potential sources of bias in observational investigations as well as strategies used to compensate for these biases, and discuss a framework to implement observational research in CER.

Conclusions: The CER paradigm recognizes the limitations of hierarchical models of evidence and favors application of a strength-of-evidence model. In this model, observational research fills gaps in randomized clinical trial data and is particularly valuable to investigate effectiveness, harms, prognosis, and infrequent outcomes as well as in circumstances where randomization is not possible and in studies of many surgical populations. Observational investigations must be designed with careful consideration of potential sources of bias and must incorporate strategies to control such bias prospectively, and their results must be reported in a uniform and transparent fashion. When these conditions can be achieved, observational research represents a valuable and critical component of modern CER.

Keywords: clinical investigation, comparative effectiveness research, evidence-based medicine, observational research, outcomes research.

Introduction

Comparative effectiveness research (CER) is defined by the Institute of Medicine (IOM) as, "the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve delivery of care [1]" and by the Federal Coordinating Council for Comparative Effectiveness Research [created by the US Congress [2]] as, "the conduct and synthesis of research comparing the benefits and harms of different interventions and strategies to prevent, diagnose, treat, and monitor health conditions in "real world" settings [3]." The purpose of CER, as stated by the IOM, is "to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both the individual and population levels [4]." CER therefore seeks to inform clinical decisions between alternate management strategies using data that reflects real patient populations and real-world clinical scenarios for the purpose of improving patient outcomes.

The potential for such a strategy to advance the public health while simultaneously decreasing health-care expenditure has captured the attention of health policymakers, and over the past several years, the CER paradigm has become part of the national discourse regarding strategies to refine the US health-care system. The connection between cost-effectiveness research and value-based medicine and CER has caused a degree of clinician and public skepticism of this model, but the overall CER paradigm reaches far beyond cost considerations. Recognizing the limitations of randomized controlled trials, CER has directed new emphasis on problem-based research and practical clinical trials as well as on observational studies that fill knowledge voids left by randomized clinical trials. Utilizing all of these modalities, CER seeks to combine treatment efficacy data with quality of life, outcomes, and other forms of effectiveness data to guide selection of optimal patient management strategies.

The concept of evidence-based medicine (EBM) is familiar to most clinicians, and often, those who encounter CER in the literature erroneously conceptualize EBM and CER as different versions of the same research paradigm. This is inaccurate, as EBM focuses primarily on demonstrating efficacy of treatment modalities, whereas CER attempts to inform clinical decisions regarding patient management. One of the most significant differences in the two research models is the role of observational research. Frequently marginalized in EBM as low-quality and suboptimal, observational research is gaining prominence as an expeditious and cost-effective modality to inform large numbers of real-world patient management decisions in the CER paradigm [4], which recognizes that there are multiple clinical situations where the unique characteristics of observational investigations can inform the decision-making process.

In this review, we highlight the role of observational research in the CER paradigm. We will review the scope and nature of observational research, present the rationale for its inclusion in CER investigations, and discuss a framework for its implementation. This discussion will also highlight the important differences between CER and EBM that are often underappreciated by clinicians.

Hierarchical Models of Evidence and CER

Discussing the role of observational research in CER requires first addressing the concept of evidence hierarchies, because one critical step in moving from an EBM to a CER paradigm is

Address correspondence to: Nicholas F. Marko, Department of Neurosurgery, S80, 9500 Euclid Avenue, Cleveland, OH 44195 USA. E-mail: markon@ccf.org

10.1111/j.1524-4733.2010.00786.x

rethinking the hierarchical scheme to stratify levels of clinical evidence upon which most EBM reviews are based. The hierarchical system utilized in the EBM model defines five levels of evidence. Randomized clinical trials (and systematic reviews thereof) are assigned to level 1, representing the highest grade of available evidence. Level 2 generally reflects cohort studies, level 3 represents case-control studies, and levels 4 and 5 are reserved for case series and expert opinions, respectively [5]. Inherent in this scheme is the belief that studies in a given level are methodologically superior to those assigned to lower levels and therefore yield higher quality, less biased results [6]. This model has gained widespread acceptance since a landmark paper in 1982 [7] demonstrated relative reductions in selection bias resulting from application of the randomized, controlled design, but recent theoretical concerns and empiric investigations have challenged the universal supremacy of the randomized clinical trials (RCTs). In so doing, they have also called into question the validity of a rigid, hierarchical model of clinical evidence.

Theoretical challenges to the hierarchical evidence model question the validity of the assumptions upon which RCTs are based. First, RCTs assume a state of clinical equipoise [8–10] that is typically attributable to a lack of sufficient scientific knowledge to definitively recommend one management strategy over another. In practice, however, the validity of this assumption may be questionable because the existing body of clinical investigations relative to a specific treatment is rarely completely neutral and true clinical equipoise may be relatively uncommon [9,11,12]. Next, the balance between the magnitude of specific and nonspecific effects of alternate treatment modalities can influence the statistical significance and thus the conclusions of RCTs, giving rise to the so-called "efficacy paradox" [11]. RCTs are typically optimized to detect differences in the magnitude of specific treatment effects while neglecting nonspecific and overall treatment effects, yet it is the latter that has been shown to be of greatest interest to patients [13–15]. Finally, the context in which the treatment is provided in RCTS as well as the external validity of the conclusions drawn by these studies must be carefully interpreted, as studies have shown that contextual therapeutic messages and beliefs can affect measured treatment efficacy [16–19] and that selection biases [20] and incongruities in study versus clinical patient populations [21,22] can affect external validity [11]. Whereas RCTs remain a critical and time-tested tool for conducting high-quality medical research, these concerns regarding the assumptions necessary for RCT validity suggest that alternate research modalities may be necessary to inform clinical decisions, particularly in clinical situations where one or more of the aforementioned preconditions for RCT validity are clearly violated.

Objective attempts to study the alleged superiority of RCTs have also called into question the position of RCTs at the top of an evidence hierarchy [23]. Multiple investigations comparing RCTs to well-designed, nonrandomized trials have consistently demonstrated high degrees of concordance in the trial results [6,24–28], demonstrating that RCTs are not necessarily superior to other modalities in their ability to produce valid clinical results. Additionally, comparisons of multiple RCTs that address similar clinical questions have demonstrated significantly discordant results [29], highlighting the fact that RCTs do not *de facto* provide definitive and incontrovertible results. Taken together, these findings suggest that designation of RCTs as the gold standard for clinical research in every and all cases may be uncertain and that alternate research designs may be equally or, at times, more appropriate to address specific clinical questions.

If the current state of clinical trial design is such that there is neither a single, gold-standard modality that fits all clinical questions nor definitive evidence of superiority among well-constructed investigations using alternate study design models, then the validity and utility of a hierarchical stratification of evidence may need to be reexamined in certain cases [6,11,30]. Several authors have argued for alternate models that inform clinical decisions by emphasizing the appropriateness of study design to the clinical question and the quality of the individual investigations rather than their hierarchical stratification [6]. This approach, sometimes described as a *circular model* of evaluation [11], has been applied in the social sciences and alternative medicine for some time [31–35]. These models acknowledge that there may be different optimal approaches to answer different clinical questions and that a composite of such investigations may provide a more comprehensive basis to inform clinical decisions than any one investigation alone.

The CER enterprise in the United States has recognized these concerns and has approached this issue by adopting a *strength-of-evidence* model rather than a *level-of-evidence* model. The United States *Agency for Health Care Research and Quality (AHRQ)* notes that, while evidence hierarchies focus primarily on study design, strength-of-evidence models focus on several additional domains to ensure methodological consistency to grade evidence and to facilitate more uniform interpretation of the meaning of evidence grading by end users of these data [36,37]. Based on evidence analysis performed by the *US Preventive Services Task Force* [38] and the *Grading of Recommendations Assessment, Development and Evaluation* working group [39], the AHRQ recommended analysis of four "required domains" when determining the strength of evidence in CER investigations: risk of bias, consistency between studies of similar questions, directness of association between interventions and outcomes of interest, and precision of evidence for anticipated outcomes [36]. Secondary domains, including dose–response associations, existence of confounders, magnitude of effect, and publication bias, can be included, when applicable, to describe the strength of evidence more robustly. A strength-of-evidence analysis results in classification of evidence into one of four grades, which describe the general magnitude of confidence that the evidence reflects a true effect. Evidence can be assigned a designation of high, moderate, or low or can be designated as insufficient to provide a conclusion [36]. More detailed guidelines for constructing and interpreting strength-of-evidence analyses have been produced by the AHRQ [36], and such analyses represent an important dimension of CER. The decision to follow a strength-of-evidence model presents an opportunity for well-designed, observational studies to assume a more prominent role in CER.

Observational Studies and CER

Definitions

The term, "observational research," has classically been used to describe investigations where the researcher observed but did not interact with the study population, and it is still often applied this way in social science [40] and business and economics research [41]. Although there remains some variability in nomenclature, the term "observational research," is typically used in modern biomedical research to refer to investigations conducted on samples whose membership has not been affected by the investigator. In this manifestation, observational research is held in contrast to randomized trial research, wherein the investigator uses a process of randomization to assign participants to specific management groups. Because such assignment does not occur in observational studies, the terms "nonrandomized research,"

"nonrandomized trials," or "nonexperimental research" are often considered to be functionally interchangeable with "observational research" [42].

Observational studies can be divided into prospective or retrospective analyses. Retrospective, observational analyses identify a sample of patients with a known outcome and seek to determine the nature and frequency of a given exposure that preceded the outcome. Similar investigation of exposure in a sample without evidence of the outcome may be used to study the outcome frequency in the general population. Members of the sample group exhibiting the outcome in question are referred to as *cases*, whereas those without the outcome of interest are considered *controls*. This designation gives rise to the term *case-control study* and describes one form of retrospective, observational research [42]. Other forms include case series and single case reports reviewed in a retrospective fashion, as well as research derived from retrospectively collected databases. Incidence rates cannot be determined from this model, but the measure of association between exposure and outcome can be investigated in case-control studies and is expressed as the *odds ratio* [43,44].

In contrast, prospective, observational analyses follow over a period of time or to a specific end point a predetermined sample of patients who have been exposed to some factor under investigation. These studies can be descriptive in nature, reporting outcomes in the group after exposure to the factor without direct, statistical comparison to a control sample, or can be analytic in nature, using comparison between the experimental sample and a control sample to calculate the probabilistic relationship between exposure and outcome [42]. The former is described as a *prospective case series,* whereas the latter are called *cohort studies* in reference to the cohort of patients that is being followed. Whereas case series are not used to calculate analytic epidemiologic statistics, cohort studies can be used to measure the *incidence rates* in the two populations as well as the *relative risk* of outcome associated with exposure [42,45].

Bias and the Limitations of Observational Studies

All studies, regardless of design, are subject to similar potential sources of bias. The Cochrane Collaboration [46], the United Kingdom's National Institute for Health and Clinical Excellence [47], and other evidence-based research bodies have organized the major sources of bias in clinical investigations into four categories. Like all study designs, observational investigations have some limitations in their ability to fully address these biases. Understanding the effects of such bias and the strategies used to limit these effects is critical to performing high-quality observational research.

Selection bias is defined as a distorted estimate of the effect that results from the way in which subjects are selected for the study population [48]. Selection bias is often attributable to the presence of *confounding variables (confounders)*, defined as extraneous variables that correlate with both the dependent and the independent variable(s). In order to be truly confounding, a variable must satisfy two conditions; it must be a risk factor for the outcome being studied and it must be associated with, but not a consequence of, the exposure being examined [49]. Closely related and commonly misunderstood is the concept of *confounding by indication*. This term describes the presence of an extraneous determinant of the outcome parameter that is present if a perceived high risk or poor prognosis is an indication for intervention [9,10,37,48,50,51]. In this circumstance, the indication correlates with the intervention and is a risk for the illness, qualifying it as a confounder and producing an imbalance in prognostic factors between comparison groups [48]. Randomization minimizes selection bias and theoretically eliminates confounders and confounding by indication, whereas the lack of randomization in observational investigations makes them particularly susceptible to these sources of bias. Control of known confounders can minimize these effects in cohort studies, and matching can address similar issues in case-control investigations [46,52]. Unknown or unmeasured confounders, however, cannot be addressed in this fashion, and the inability of observational studies to address such bias represents the most significant limitation of this study design.

Performance bias refers to systematic differences in care other than intervention under investigation that is provided to study participants, whereas *detection bias* describes systematic differences between comparison groups in outcomes assessments [46]. Both of these biases are addressed in RCTs by blinding, but complete blinding is usually not possible in observational research. Performance bias is addressed in observational studies, including both cohort and case-control studies, by measuring the exposure to the intervention of interest and ensuring that differences in the exposure of the comparison groups either did not exist or did not affect the investigational outcomes. Cohort studies can implement limited blinding strategies to reduce detection bias, whereas case-control studies must rely on the case definition to reduce this bias [46,47,52]. Limitations on blinding and the resulting reduced control of performance and detection biases represent additional limitations of observational studies. *Attrition bias* refers to systematic differences in the loss of participants between the comparison groups [46,52]. Attrition has significant potential to affect all types of clinical trials, regardless of design. Close and meticulous follow-up is the best strategy to limit the effects of attrition bias in both randomized controlled trials and in observational research [46,52].

Advantages of Observational Studies

Despite having some vulnerability to the aforementioned sources of potential bias, observational studies are well suited to investigate certain types of clinical questions that cannot be adequately addressed by RCTs alone [53,54]. Several examples of clinical situations where observational investigations can provide clinical relevant data beyond that afforded by traditional RCTs are as follows:

Investigations of effectiveness. RCTs are designed for the specific purpose of evaluating the *efficacy* of a particular treatment, or the extent to which the treatment under study produces the intended effect in an ideal setting. Such studies use homogeneous populations, under carefully controlled conditions, to study a limited number of immediate or short-term outcomes. *Effectiveness* studies, which are central to CER, have the more complex goal of assessing the beneficial effects of the particular treatment in the context of everyday practice [39]. Here, the patient population is more heterogeneous, conditions are less controlled, and multiple variables may impact the outcomes. Studying the beneficial effects of therapy therefore frequently requires larger sample populations that are followed over longer periods of time, a situation well suited for observational research (particularly prospective trials and patient databases). Additionally, effectiveness research requires that the effects of a number of potential variables and confounders be analyzed rather than eliminated. This includes variables introduced by the complex decision-making processes of actual caregivers, which may encompass any of a number of factors that motivate decisions regarding selection, initiation, adjustment, or cessation of therapy [55]. Such

scenarios require the type of comprehensive data collection that is characteristic of observational research. Ultimately, effectiveness trials attempt to assess the overall degree of benefit in "real world," clinical settings and to improve the *generalizability*, or external validity, of RCTs [30,55–58]. This is central to the mission of CER and is a major reason why observational research has an important role in this paradigm.

Investigation of harms. If effectiveness research is conceptualized as focusing on the real-world benefits of a particular treatment, then *harms* investigations can be conceived as the inverse, focusing on the potential negative consequences of the treatment [59]. Like effectiveness research, reliance on RCTs for harms investigations has several fundamental limitations. In addition to considerations of sample size [60] and homogeneity [61], duration of study [60], and ability to examine multiple variables [10], RCTs are further limited in harms investigation because they lack specific hypotheses designed to identify adverse events [62]. Instead, identification of such harms represents a secondary consideration in most RCTs. They are therefore either underpowered for or biased against harms detection by virtue of their design, which results in inadequate or incomplete harms reporting [63–66]. Finally, RCTs do not generally compare the results of alternate treatment strategies [67,68], so information regarding relative harms is almost never available from such investigations. In contrast, observational studies are well suited to address all of these concerns [9,10,30,59]. Sample size, homogeneity, and duration of study have been addressed for effectiveness studies, and similar reasoning is applicable to the advantages of observational research in detection of harms. Additionally, observational studies can be designed to identify specifically and in detail the spectrum of novel harms that may be associated with a particular treatment or to further investigate the relative frequency of known harms [59]. Finally, observational designs are particularly well suited to hypothesis testing regarding differential risk of specific harms associated with alternate interventions [59]. All of these factors are essential to CER and represent another valuable contribution of observational research.

An example of the important contribution of observational studies in assessing potential harms is the identification of the increased risk of myocardial infarction associated with the drug, rofecoxib (Vioxx®). After a 6-month priority review completed in May of 1999, rofecoxib became the second cyclooxygenase-2 inhibitor approved for symptomatic treatment of osteoarthritis [69]. The Vioxx Gastrointestinal Outcomes Research study of 2000 [70] identified a potentially increased risk of myocardial infarction in patients taking rofecoxib. This was attributed by the manufacturer (Merck, Whitehouse Station, NJ) to a cardioprotective effect of the control drug, naproxen, a claim that was ultimately refuted using a meta-analysis of 11 observational investigations [71]. Observational research played an even more significant role when the Federal Drug Administration's (FDA's) Dr. David Graham lead a collaborative effort with Kaiser Permanente of California to perform a retrospective review of cardiac complications in patients taking rofecoxib [69]. The results of this nested case-control study of 2,302,029 person-years of follow-up data demonstrated an increased risk of "serious coronary heart disease" in users of rofecoxib (odds ratio 1.59) [72] and led to Merck's voluntary withdrawal of rofecoxib in 2004 [69].

Investigations of delayed or infrequent harms. Although these studies can be considered a subtype of harms investigations, they merit brief, individual attention because they represent a circumstance where all of the aforementioned benefits of observational research (in both effectiveness and harms analyses) are simultaneously realized. Here, the size, heterogeneity, and time limitations of RCTs combine with their lack of harms hypotheses and comparative effectiveness abilities to produce a situation where data on infrequent harms would be essentially impossible to collect. In practice, it is estimated that effects that take longer than 1 year to develop or harms that occur at a rate less than 1 per 200 per year are not adequately detected by RCTs [73]. Observational research, either prospective or retrospective, can be applied to large, heterogeneous groups over long periods of time to identify such delayed or infrequent harms [74–76]. Examples of the use of observational research to identified delayed outcomes include identification of the link between diethylstilbestrol treatment and clear cell vaginal carcinoma [77–79], and observational analysis has provided information regarding the infrequent but significant associations between childhood aspirin use and Reye's syndrome [54,80,81] and between hypertrophic cardiomyopathy and sudden cardiac death in young athletes [54,82].

Investigations of prognosis and long-term or infrequent outcomes. Similar logic can be applied to studies of delayed or infrequent outcomes. Because of time and sample size limitations, respectively, RCTs are not well suited to investigating such outcomes. Accordingly, identification of long-term benefits or infrequent outcomes is difficult using the RCT model. Observational studies are more appropriate to answer these questions [74,75]. Long-term prognosis can be considered a more general category of delayed outcome given an exposure, and similar logic can be applied. Numerous outcomes studies have been performed using this model, particularly those that involve the surgical management of uncommon conditions [83] or the durability of surgical repairs or constructs [84].

Investigations without randomization. Randomization minimizes selection bias and is fundamental to the design of RCTs, but there are four situations where randomization cannot or need not be applied. Observational research, however, can be used in all four situations. First, there are circumstances where randomization is *not necessary* [30]. This generally occurs when two conditions are met; the effect under study is dramatic and the chance of confounding factors is negligible. In this situation, a well-constructed, observational study is sufficient to demonstrate effectiveness [74,75]. Second, there are situations where randomization may be considered *impossible*. Several factors can lead to this situation, including refusal by clinicians or patients to participate in a proposed trial, political or legal obstacles to randomization, or logistical issues relating to cost or availability of resources [74]. In all of these situations, observational research may remain a viable alternative. Third, there are circumstances where randomization into a clinical trial is considered *unethical* [8,85,86]. This may include questions of true equipoise, issues of clinician or researcher bias, considerations of patient safety or autonomy, and myriad other potential ethical dilemmas [74,75]. When randomization has been determined to be unethical, observational studies may be viable options for securing the best possible clinical evidence within the confines of the existing ethical limitations. Fourth, there are circumstances where randomization would prove to be *self-defeating*. In trials where the effectiveness of the intervention depends on the subject's active participation in or knowledge of the treatment, randomization either becomes impossible or biases the results. Observational research allows for subject participation or knowledge while retaining the ability to measure outcomes [74]. All of these "real-world" situations may

arise in CER, and observational studies are the appropriate tools to address these clinical circumstances.

Investigations in surgical populations. Patients undergoing surgical procedures represent a unique clinical circumstance, and investigations of these patients have classically favored observational research. Historically the indications for surgery have been dogmatic in some cases, as have the specific management strategies implemented by the surgeon [87]. This has drawn criticism from advocates of RCTs, who argue that randomized trials should occupy a more prominent role in surgical decision-making [87]. Although surgeons are beginning to embrace more objective outcomes research, several practical counterarguments have been made against requiring routine use of RCTs in surgical specialties. First, although basic surgical procedures may fit into an RCT model, complex or potentially hazardous interventions with serious immediate and long-term consequences have made surgeons and patients [88] wary of the role of RCTs and reluctant to accept random assignment of treatment [75]. Next, many conditions managed surgically are either relatively rare or are highly individual, reflecting the unique anatomic and physiologic consequences of the patient's disease. The same is true for the adverse outcomes of surgical procedures, which are highly dependent upon the individual characteristics of the patient, his pathology, and the surgeon's judgment and skill. RCTs are not typically designed to identify rare events that are not easily classified (see above) and therefore may not be the ideal modality to investigate many surgical procedures [88,89]. Third, the combination of long historical experience with surgical management of common conditions combined with the self-evident, prompt, and often dramatic nature of surgical results may argue against the necessity of RCTs for many surgical conditions [87]. Fourth, the logistics and cost associated with conducting surgical RCTs may be so pronounced that such studies become practically impossible [87,88,90]. Because of these circumstances, it has been estimated that only 40% of surgical treatment questions would be amenable to investigations with RCTs [88] and that only 3% to 9% of currently published surgical investigations utilize this research model [90]. As a result of a combination of patient preference, surgeon's attitudes and training, and the logical arguments outlined herein, it is likely that observational research will remain a mainstay of CER in the surgical population for the foreseeable future.

Investigations using patient registries and databases. Patient registries prospectively compile data related to patient outcomes after particular exposures, treatments, or procedures. These datasets are typically comprehensive, including data compiled from medical records, direct patient interactions, and long-term follow-up. In addition, they are often supplemented with or linked to additional administrative information, including detailed demographic information, data on proxy outcomes, and measurements of health-care resource utilization. More recently, such databases have also begun to include information on performance outcomes and on quality of life. These databases are usually constructed for specific research purposes, but their application can extend well beyond their original, intended purpose. They serve as valuable repositories of comprehensive clinical information and are considered among the strongest sources of data for observational research [9,59]. Examples of the value of observational, database research include characterization of the side effects of prostate surgery [91,92] and identification of the risk of upper gastrointestinal events associated with bisphosphonates [93].

Additional investigations. The aforementioned examples describe clinical situations where observational investigations are either superior to RCTs or provide valuable, supplemental clinical evidence. Additional applications of observational research include investigations of models for diagnosis [37,73], investigations designed to clarify outcomes of RCTs, development of strategies to identify research priorities based upon current clinical knowledge, and applications in health-care policy [75]. The details of these applications are beyond the scope of this review, but the ultimate message is that there are multiple clinical situations where the unique characteristics of observational investigations can inform medical decision-making. Vandenbroucke summarizes this view well: "when making medicine singularly dependent on randomized trials . . . I am not certain that the best interest of the patient is served. Moreover, by taking away credibility from case-control research . . . the patient and the physician are left defenseless a second time [94]."

Reporting and Quality Assessment Tools for Observational Studies

Observational studies are susceptible to multiple potential sources of bias and confounding, so concerns regarding the quality, validity, and reliability of data gathered from such studies are both abundant and well-founded. As the number of observational studies reported in the literature continues to grow, the need for both authors and consumers of health-care information to determine the quality of observational investigations becomes progressively more apparent. This is particularly important in today's research environment, where evolving trends toward aggregating results of empiric studies into systematic reviews and meta-analyses have lead to an emerging body of such literature that includes observational studies [95]. A study measuring the magnitude of this effect noted that from 1955 to 1992, there were a total of 678 published meta-analyses of observational studies, while 525 were reported from 1992 to 1995, and 400 were reported in 1996 alone [96].

Proponents of including observational data in meta-analyses and systematic reviews cite many of the strengths of observational data discussed above as compelling reasons for the inclusion of such data in evidence that informs medical decisions. This position is supported by investigations comparing meta-analyses based on RCTs to those based on "high-quality" observational studies, which demonstrate similar estimates of effect from both [24,25,97,98]. Critics of inclusion of observational data in meta-analyses and systematic reviews cite evidence of compounding effects of bias and confounding that can lead to potentially distorted conclusions in such analyses [95,99]. In an effort to reconcile these positions and to ensure reporting of reliable data, a considerable amount of attention has focused on developing and validating tools to assess the quality of observational investigations [100–103].

Before discussing tools for quality assessment in observational research, it is important to define the concept of quality. Although it can be an amorphous concept, "quality" typically refers to "susceptibility to bias [101]." This definition notwithstanding, many quality metrics include considerations of factors not directly related to bias, including sample size, power calculations, and ethical considerations. Additionally, it is important to distinguish between performance quality and reporting quality. The former refers to study design and implementation that limits bias, whereas the latter refers to adherence to comprehensive reporting standards that make research transparent but do not necessarily modulate the effects of bias [101,104]. Efforts to improve both dimensions of quality in observational

studies have been well described but remain an area of active investigation.

Tools to assess the quality of performance, or susceptibility to bias, often take the form of numeric scales, checklists, or checklists with a summary judgment [101]. A recent review of such metrics for observational research identified 86 tools for quality of performance assessment [101], although the number of such metrics continues to rise. These tools assess selection methods, measurement of study variables, design-specific sources of bias, control of confounders, and application of statistical methods. The design process, scope, and generalizability were variable among these tools. Approximately half of the quality metrics assessed what are considered to be the three fundamental quality assessment domains, including patient selection, appropriate measurement of variables, and appropriate control of bias and confounding [101]. Overall, there is a belief that authors of observational studies should use some form of quality of performance assessment tool, although there has been no consensus on a single tool that most reliably achieves this goal [101,103].

Quality of reporting must also be assessed when interpreting observational research. Investigations regarding this dimension of quality have demonstrated that important information on study design is often not reported in observational literature, including eligibility criteria [103], methods used to identify the study and control populations [105], and rationale behind the choice of potential confounding variables [106]. Omission of these data makes critical interpretation of this literature difficult [102], which may result in delayed realization of the benefits of the investigations [107]. Based upon these findings and utilizing consensus models adopted for reporting systematic reviews (QUOROM) [108], randomized trials (CONSORT) [52,109,110], and studies of diagnostic tests (STARD) [111], quality of reporting standards have been proposed for observational investigations. The *Strengthening the Reporting of Observational Studies in Epidemiology* consensus statement [102] produced a checklist of 22 items that outline specific reporting methods and standards for observational studies. These items include detailed recommendations regarding reporting of the title, abstract, introduction, methods, results, discussion, and additional information in observational research [102]. Similar efforts to standardize reporting of meta-analyses of observational investigations have been made by the *Meta-analysis of Observational Studies in Epidemiology* (MOOSE) group [100], who generated a 35-item checklist covering reporting in the domains of background, search strategy, methods, results, discussion, and conclusions for such meta-analyses [100]. The extent to which these recommendations will be adopted by the scientific community remains to be seen, but these efforts demonstrate the emerging emphasis placed on quality of performance and reporting in observational investigations.

Applying Observational Investigation to CER

The preceding discussion has reviewed the strengths and limitations of observational studies and has highlighted the potential role for such investigations in patient-centered CER. Successful use of observational studies in CER, however, requires selection of appropriate clinical questions to which this methodology should be applied and, subsequently, application of a logical framework for implementing the research plan. This issue has been studied independently by the AHRQ [9,10] and the investigators of the *Good Research for Comparative Effectiveness (GRACE)* initiative [55]. Both agencies have issued guidelines for implementation of observational investigations in CER, which are summarized as follows:

I. *Utilize a logical approach to identify evidence gaps where observational studies may be particularly applicable*, including situations where:
   a. *Determinants of use are not related to determinants of outcome:* This often occurs when treatment decisions are primarily affected by reimbursement patterns, patients' insurance coverage, or drug formularies. Bias is reduced when experimental groups are divided along these lines because there is theoretically little influence between group membership and clinical outcomes.
   b. *Determinants of treatment are inconsistent or unknown:* This occurs in situations of clinical equipoise or when there is a pure lack of appropriate evidence to inform clinical decisions. In these circumstances, treatment is largely dictated by clinician preference, and outcomes from different management strategies can be readily compared.
   c. *Treatment decisions are made based upon factors unrelated to the outcomes of interest:* When treatment is initiated, adjusted, or discontinued because of a factor unrelated to the outcome of interest, distinct clinical groups can be segregated based upon a factor not related to the outcome under study.
   d. *Little relevant information is available and there is need of preliminary evidence:* In this situation, observational investigations may be reasonable initial studies to uncover some fundamental information related to the clinical question at hand, which can then be used to guide more focused investigation.

II. *Prepare a study plan in advance*. High-quality observational research typically results from careful planning and meticulous study design. This represents an aggressive attempt to identify and control bias and confounders by:
   a. Defining the hypothesis and purpose;
   b. Identifying the populations, treatments, and comparators;
   c. Defining the outcomes that will be used to measure effectiveness of each treatment;
   d. Considering the necessary sample size and performing power calculations.

III. *Collect clinically relevant data as efficiently as possible*. In preparation for and during the process of data collection, it is necessary to evaluate and reevaluate the following:
   a. Inclusion criteria and enrollment;
   b. The specific data that will be (are being) collected and the checks in place to assure validity of this data;
   c. Potentially useful data that are not being captured.

IV. *Analyze data using a patient-centered model*. Here, it is important to ensure that the data being analyzed and the resultant conclusions reflect "real-world" circumstances and are patient-centric. This involves asking questions, such as:
   a. Do the study results reflect outcomes of actual treatment that has been delivered? As opposed to intention-to-treat analyses, CER must study treatments that have actually been administered.
   b. Have people with similar clinical characteristics, including disease severity and opportunities for treatment, been studied?

c. How well have alternate explanations for the findings been considered and how carefully have potential biases been investigated and controlled?

This model, which has been forwarded primarily by the GRACE initiative [55], specifically highlights fundamental issues of good study design as applied to observational studies in CER. Although the exact model for implementing observational research in CER studies will vary with the particular clinical question and research environment, the central principle remains that such research will be most valuable when it is used to address gaps in the existing literature, is targeted to clinical situations for which observational research is appropriate based upon considerations of its strengths and limitations, and remains patient-centered.

Conclusions

The comparative effectiveness (CER) paradigm recognizes the limitations of hierarchical models of evidence, particularly in patient populations that are not well represented in RCT design and in clinical circumstances where the necessary preconditions for RCT validity are not met. Instead, CER favors application of a strength-of-evidence model where observational investigations occupy an important role to fill knowledge voids left by RCTs. Observational research is particularly valuable to investigate effectiveness, harms, prognosis, and infrequent outcomes as well as in circumstances where randomization is not possible and in many studies of some surgical populations. Observational investigations must be designed with careful consideration of potential sources of bias and must incorporate strategies to control such bias prospectively, and their results must be reported in a uniform and transparent fashion. When these conditions can be achieved, observational research represents a valuable and critical component of modern CER.

Source of financial support: None.

References

1 Initial national priorities for comparative effectiveness research: report brief. Institutes of Medicine of the National Academies. 2009.
2 Public Law 111-5. The American Recovery and Reinvestment Act. 2009.
3 Federal Coordinating Council for Comparative Effectiveness Research: Report to the President and the Congress. Washington, DC: US Department of Health and Human Services, June 30, 2009.
4 Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. Ann Intern Med 2009;151:203–5.
5 Oxford Centre for Evidence-based Medicine Levels of Evidence (March 2009). Centre for Evidence Based Medicine. 2009.
6 Concato J. Observational versus experimental studies: what's the evidence for a hierarchy? NeuroRx 2004;1:341–7.
7 Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. Am J Med 1982;72:233–40.
8 Lilford RJ, Jackson J. Equipoise and the ethics of randomization. J R Soc Med 1995;88:552–9.
9 Gliklich R, Dreyer N, eds. Registries for Evaluating Patient Outcomes: A User's Guide (Prepared by Outcome DEcIDE Center [Outcome Sciences, Inc. dba Outcome] under Contract No. HHSA29020050035ITO1.) AHRQ Publication No. 07-EHC0001-1. Rockville, MD: Agency for Healthcare Research and Quality, 2007.
10 AHRQ. Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews. Version 1.0. Rockville, MD: AHRQ, 2007.
11 Walach H, Falkenberg T, Fonnebo V, et al. Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. BMC Med Res Methodol 2006;6:29.
12 Lefering R, Neugebauer E. Problems of randomized controlled trials (RCT) in surgery. In: Abel U, Koch A, eds. Nonrandomized Comparative Clinical Studies. Dusseldorf: Symposion Publishing, 1998.
13 Barrett B, Marchand L, Scheder J, et al. Bridging the gap between conventional and alternative medicine. J Fam Pract 2000;49:234–9.
14 Moore J, Phipps K, Marcer D, et al. Why do people seek treatment by alternative medicine? Br Med J (Clin Res Ed) 1985;290:28–9.
15 Lewith GT, Bensoussan A. Complementary and alternative medicine—with a difference. Med J Aust 2004;180:585–6.
16 Kleijnen J, de Craen AJ, van Everdingen J, et al. Placebo effect in double-blind clinical trials: a review of interactions with medications. Lancet 1994;344:1347–9.
17 Khan A, Khan S. Placebo in mood disorders: the tail that wags the dog. Curr Opin Psychiatry 2003;16:35–9.
18 Kirsch I. Response expectancy as a determinant of experience and behavior. Am Psychol 1985;40:1189–202.
19 Walach H, Kirsch I. Herbal treatments and antidepressant medication: Similar data, divergent conclusions. In: Lilienfeld S, Lynn S, Lohr J, eds. Science and Pseudoscience in Clinical Psychology. New York: Guilford, 2003.
20 Patten SB. Selection bias in studies of major depression using clinical subjects. J Clin Epidemiol 2000;53:351–7.
21 Feinstein A. Problems of randomized trials. In: Abel U, Koch A, eds. Nonrandomized Comparative Clinical Studies. Dusseldorf: Symposion Publishing, 1998.
22 Wragg JA, Robinson EJ, Lilford RJ. Information presentation and decisions to enter clinical trials: a hypothetical trial of hormone replacement therapy. Soc Sci Med 2000;51:453–62.
23 Bigby M. Challenges to the hierarchy of evidence: does the emperor have no clothes? Arch Dermatol 2001;137:345–6.
24 Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. N Engl J Med 2000;342:1878–86.
25 Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med 2000;342:1887–92.
26 McKee M, Britton A, Black N, et al. Methods in health services research. Interpreting the evidence: choosing between randomised and non-randomised studies. BMJ 1999;319:312–5.
27 Horwitz RI, Viscoli CM, Clemens JD, et al. Developing improved observational methods for evaluating therapeutic effectiveness. Am J Med 1990;89:630–8.
28 A randomized trial of propranolol in patients with acute myocardial infarction. I. Mortality results. JAMA 1982;247:1707–14.
29 Horwitz RI. Complexity and contradiction in clinical trial research. Am J Med 1987;82:498–510.
30 Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. BMJ 2004;328:39–41.
31 Shadish W, Cokk T, Leviton L. Foundations of Program Evaluation. Theories of Practice. Newbury Park, CA: Sage, 1991.
32 Chen H, Rossi P. Evaluating with sense. The theory-driven approach. Eval Rev 1983;7:283–302.
33 Cook T, Wittman W. Lessons learned about evaluation in the United States and some possible implications for Europe. Eur J Psychol Assess 1998;14:97–115.
34 Rossi P, Freeman H. Evaluation. A Systematic Approach (2nd ed.). Beverly Hills: Sage, 1982.
35 Wittman W, Walach H. Evaluating complementary medicine: Lessons to be learned from evaluation research. In: Lweith G, Jonas W, Walach H, eds. Clinical Rsearch in Complementary Therapies: Principles, Problems, and Solutions. London: Churchill Livingston, 2002.
36 Owens D, Lohr K, Atkins D, et al. Grading the strength of a body of evicence when comparing medical interventions. Rockville, MD, 2009.
37 Vandenbroucke JP. Observational research, randomised trials, and two views of medical science. PLoS Med 2008;5:e67.

38 Harris RP, Helfand M, Woolf SH, et al. Current methods of the US Preventive Services Task Force: a review of the process. Am J Prev Med 2001;20:21–35.
39 Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924–6.
40 Lewis-Beck MS, Bryman A, Liao TF. The Sage Encyclopedia of Social Science Research Methods. Thousand Oaks, CA: Sage, 2004.
41 Abrams B, American Marketing Association. The Observational Research Handbook : Understanding How Consumers Live with Your Product. Lincolnwood, IL: NTC Business Books, 2000.
42 Rothman KJ, Greenland S, Lash TL. Modern Epidemiology (3rd ed.). Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008.
43 Pearce N. What does the odds ratio estimate in a case-control study? Int J Epidemiol 1993;22:1189–92.
44 Cummings P. The relative merits of risk ratios and odds ratios. Arch Pediatr Adolesc Med 2009;163:438–45.
45 McNutt LA, Wu C, Xue X, et al. Estimating the relative risk in cohort studies and clinical trials of common outcomes. Am J Epidemiol 2003;157:940–3.
46 Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 [Updated September 2006]. The Cochrane Library, Issue 4. Chichester: John Wiley & Sons, Ltd., 2006.
47 National Institute for Health and Clinical Excellence. The Guidelines Manual. London: National Institute for Health and Clinical Excellence, 2009.
48 Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. Am J Epidemiol 1999;149:981–3.
49 Schlesselman J. Case-Control Studies. Design, Conduct, Analysis. New York: Oxford University Press, 1982.
50 Miettinen O. Theoretical Epidemiology: Principles of Occurrence Research in Medicine. New York: John Wiley & Sons, 1985.
51 Shrier I, Boivin JF, Steele RJ, et al. Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. Am J Epidemiol 2007;166:1203–9.
52 Deeks JJ, Dinnes J, D'Amico R, et al. Evaluating non-randomised intervention studies. Health Technol Assess 2003;7:iii-x,1–173.
53 Chou R, Helfand M. Challenges in systematic reviews that assess treatment harms. Ann Intern Med 2005;142:1090–9.
54 Ray JG. Evidence in upheaval: incorporating observational data into clinical practice. Arch Intern Med 2002;162:249–54.
55 Dreyer N. GRACE principles: good research for comparative effectiveness. 2008.
56 Helfand M, Balshem H. Principles in developing and applying guidance: AHRQ and the effective healthcare program. J Clin Epidemiol 2010;63:484–90.
57 Avorn J. In defense of pharmacoepidemiology—embracing the yin and yang of drug research. N Engl J Med 2007;357:2219–21.
58 Godwin M, Ruhland L, Casson I, et al. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. BMC Med Res Methodol 2003;3:28.
59 Chou R, Aronson N, Atkins D, et al. Assessing harms when comparing medical interventions: AHRQ and the Effective Health-care Program. J Clin Epidemiol 2008;63:502–12.
60 Vandenbroucke JP. Benefits and harms of drug treatments. BMJ 2004;329:2–3.
61 Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?" Lancet 2005;365:82–93.
62 Ioannidis JP, Evans SJ, Gotzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med 2004;141:781–8.
63 Edwards JE, McQuay HJ, Moore RA, et al. Reporting of adverse effects in clinical trials should be improved: lessons from acute postoperative pain. J Pain Symptom Manage 1999;18:427–37.
64 Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. JAMA 2001;285:437–43.
65 Loke YK, Derry S. Reporting of adverse drug reactions in randomised controlled trials—a systematic survey. BMC Clin Pharmacol 2001;1:3.
66 Papanikolaou PN, Churchill R, Wahlbeck K, et al. Safety reporting in randomized trials of mental health interventions. Am J Psychiatry 2004;161:1692–7.
67 Chou R, Fu R, Huffman LH, et al. Initial highly-active antiretroviral therapy with a protease inhibitor versus a non-nucleoside reverse transcriptase inhibitor: discrepancies between direct and indirect meta-analyses. Lancet 2006;368:1503–15.
68 Song F, Altman DG, Glenny AM, et al. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. BMJ 2003;326:472.
69 Kweder S. Vioxx and Drug Safety: Statement to the United States Senate Finance Committee, 2004.
70 Bombardier C, Laine L, Reicin A, et al. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group. N Engl J Med 2000;343:1520–8. 2 p following 28.
71 Juni P, Nartey L, Reichenbach S, et al. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. Lancet 2004;364:2021–9.
72 Graham DJ, Campen D, Hui R, et al. Risk of acute myocardial infarction and sudden cardiac death in patients treated with cyclo-oxygenase 2 selective and non-selective non-steroidal anti-inflammatory drugs: nested case-control study. Lancet 2005;365:475–81.
73 Vandenbroucke JP. When are observational studies as credible as randomised trials? Lancet 2004;363:1728–31.
74 Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ 1996;312:1215–8.
75 Norris SL, Atkins D. Challenges in using nonrandomized studies in systematic reviews of treatment interventions. Ann Intern Med 2005;142:1112–9.
76 Jager KJ, Stel VS, Wanner C, et al. The valuable contribution of observational studies to nephrology. Kidney Int 2007;72:671–5.
77 Veurink M, Koster M, Berg LT. The history of DES, lessons to be learned. Pharm World Sci 2005;27:139–43.
78 Hill EC. Clear cell carcinoma of the cervix and vagina in young women. A report of six cases with association of maternal stilbestrol therapy and adenosis of the vagina. Am J Obstet Gynecol 1973;16:470–84.
79 Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. N Engl J Med 1971;284:878–81.
80 Starko KM, Ray CG, Dominguez LB, et al. Reye's syndrome and salicylate use. Pediatrics 1980;66:859–64.
81 Waldman RJ, Hall WN, McGee H, et al. Aspirin as a risk factor in Reye's syndrome. JAMA 1982;247:3089–94.
82 Maron BJ, Gohman TE, Aeppli D. Prevalence of sudden cardiac death during competitive sports activities in Minnesota high school athletes. J Am Coll Cardiol 1998;32:1881–4.
83 Talamonti G, D'Aliberti G, Collice M. Myelomeningocele: long-term neurosurgical treatment and follow-up in 202 patients. J Neurosurg 2007;107:368–86.
84 Stauffer RN. Ten-year follow-up study of total hip replacement. J Bone Joint Surg Am 1982;64:983–90.
85 Kodish E, Lantos JD, Siegler M. The ethics of randomization. CA Cancer J Clin 1991;41:180–6.
86 Retsas S. Treatment at random: the ultimate science or the betrayal of Hippocrates? J Clin Oncol 2004;22:5005–8; discussion 09–11.
87 Meakins JL. Innovation in surgery: the rules of evidence. Am J Surg 2002;183:399–405.
88 Solomon MJ, McLeod RS. Should we be performing more randomized controlled trials evaluating surgical operations? Surgery 1995;118:459–67.

89 Hartling L, McAlister FA, Rowe BH, et al. Challenges in systematic reviews of therapeutic devices and procedures. Ann Intern Med 2005;142:1100–11.
90 Solomon MJ, McLeod RS. Surgery and the randomised controlled trial: past, present and future. Med J Aust 1998;169:380–3.
91 Fowler FJ Jr, Barry MJ, Lu-Yao G, et al. Patient-reported complications and follow-up treatment after radical prostatectomy. The national medicare experience: 1988–1990 (updated June 1993). Urology 1993;42:622–9.
92 Wennberg JE, Roos N, Sola L, et al. Use of claims data systems to evaluate health care outcomes. Mortality and reoperation following prostatectomy. JAMA 1987;257:933–6.
93 Ettinger B, Pressman A, Schein J. Clinic visits and hospital admissions for care of acid-related upper gastrointestinal disorders in women using alendronate for osteoporosis. Am J Manag Care 1998;4:1377–82.
94 Vandenbroucke JP. Observational research and evidence-based medicine: what should we teach young physicians? J Clin Epidemiol 1998;51:467–72.
95 Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. BMJ 1998;316:140–4.
96 Thacker S, Stroup D, Olsen C. Characteristics of Meta-Analyses Submitted to A General Medical Journal. Syst Rev Evid Action Int Cochrane Colloq 6th. Baltimore, MD: Centers for Disease Control and Prevention, 1998.
97 Furlan AD, Tomlinson G, Jadad AA, et al. Methodological quality and homogeneity influenced agreement between randomized trials and nonrandomized studies of the same intervention for back pain. J Clin Epidemiol 2008;61:209–31.
98 MacLehose RR, Reeves BC, Harvey IM, et al. A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies. Health Technol Assess 2000;4:1–154.
99 Reeves BC, van Binsbergen J, van Weel C. Systematic reviews incorporating evidence from nonrandomized study designs: reasons for caution when estimating health effects. Eur J Clin Nutr 2005;59(Suppl. 1):S155–61.
100 Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of observational studies in epidemiology (MOOSE) group. JAMA 2000;283:2008–12.
101 Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. Int J Epidemiol 2007;36:666–76.
102 von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med 2007;4:e296.
103 Tooth L, Ware R, Bain C, et al. Quality of reporting of observational longitudinal research. Am J Epidemiol 2005;161:280–8.
104 Huwiler-Muntener K, Juni P, Junker C, et al. Quality of reporting of randomized trials as a measure of methodologic quality. JAMA 2002;287:2801–4.
105 Lee W, Bindman J, Ford T, et al. Bias in psychiatric case-control studies: literature survey. Br J Psychiatry 2007;190:204–9.
106 Pocock SJ, Collier TJ, Dandreo KJ, et al. Issues in the reporting of epidemiological studies: a survey of recent practice. BMJ 2004;329:883.
107 Bogardus ST Jr, Concato J, Feinstein AR. Clinical epidemiological quality in molecular genetic research: the need for methodological standards. JAMA 1999;281:1919–26.
108 Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of reporting of meta-analyses. Lancet 1999;354:1896–900.
109 Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996;276:637–9.
110 Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. JAMA 2001;285:1987–91.
111 West S, King V, Carey TS, et al. Systems to rate the strength of scientific evidence. Evid Rep Technol Assess (Summ) 2002;47:1–11.